

Épreuve 3 : Pandas

Module IDS - IMT Lille Douai

On dispose d'un jeu de données MovieLens¹ contenant des évaluations de films fournies par des spectateurs. L'archive `ml-1m.zip` contient les fichiers suivants :

- `movies.dat`, `ratings.dat` et `users.dat`, qui contiennent les données ;
- `README`, qui contient le descriptif des données.

À l'aide de la bibliothèque Pandas, calculez les résultats ci-dessous. Entrez vos résultats dans le formulaire disponible sur Whippet, qui vous indiquera si vos résultats sont corrects.

Défi : essayez de trouver un maximum de solutions en une seule ligne de code.

1. Le nombre total d'évaluations disponibles.
2. Le nombre total de films disponibles.
3. La note moyenne des évaluations.
4. Le nombre d'évaluations du film d'identifiant `3654`.
5. La note moyenne obtenue par le film d'identifiant `111`.
6. Le nombre d'évaluations avec une note supérieure ou égale à 4.
7. Le nombre d'évaluations avec une note inférieure ou égale à 2.
8. Le nombre d'évaluations du film d'identifiant `2021` supérieures à 3.
9. Le nombre de films dont les catégories incluent `Animation`.
10. Le nombre de films dont au moins une évaluation est supérieure ou égale à 4.
11. Le nombre de films dont toutes les évaluations sont supérieures ou égales à 4.
12. La note moyenne obtenue par le film de titre `Big Lebowski, The` (1998).
13. Le nombre d'évaluations produites par les utilisateurs de la classe d'âge `56+`.
14. La classe d'âge dont les évaluations sont les moins élevées en moyenne.
15. L'identifiant du film qui a reçu le plus d'évaluations par la classe d'âge `18-24`.

1. <http://files.grouplens.org/datasets/movielens/ml-1m.zip>